

Provenance and collections stewardship attribution standards

Dave Dubin

September 19, 2017

Review of P9 discussions in April

- The April 6 breakout session for the TDWG/RDA WG on Metadata standards for collection stewardship attribution focused on representing **role** attribution for the maintenance, curation, and digitization of collections.

Review of P9 discussions in April

- The April 6 breakout session for the TDWG/RDA WG on Metadata standards for collection stewardship attribution focused on representing **role** attribution for the maintenance, curation, and digitization of collections.
- Two specifications served as foci for these discussions:

Review of P9 discussions in April

- The April 6 breakout session for the TDWG/RDA WG on Metadata standards for collection stewardship attribution focused on representing **role** attribution for the maintenance, curation, and digitization of collections.
- Two specifications served as foci for these discussions:
 - OpenRIF/VIVO-ISF Ontology: Contributor roles, activities, and relationships in clinical research.

Review of P9 discussions in April

- The April 6 breakout session for the TDWG/RDA WG on Metadata standards for collection stewardship attribution focused on representing **role** attribution for the maintenance, curation, and digitization of collections.
- Two specifications served as foci for these discussions:
 - OpenRIF/VIVO-ISF Ontology: Contributor roles, activities, and relationships in clinical research.
 - CRediT (Contributor Roles Taxonomy) represents the diverse roles of contributors to scientific scholarly output.

Review of P9 discussions in April

- The April 7 breakout session for the Research Data Provenance Interest Group was dedicated to the proposal of a new Working Group on Provenance Patterns.

Review of P9 discussions in April

- The April 7 breakout session for the Research Data Provenance Interest Group was dedicated to the proposal of a new Working Group on Provenance Patterns.
- The main goal of this group is to identify and document general provenance solution patterns, abstracted from specific best practice solutions.

Review of P9 discussions in April

- The April 7 breakout session for the Research Data Provenance Interest Group was dedicated to the proposal of a new Working Group on Provenance Patterns.
- The main goal of this group is to identify and document general provenance solution patterns, abstracted from specific best practice solutions.
- WGPP aspires to help groups like this one recognize when solutions developed in the context of a particular use case are exemplars of more general patterns.

Review of P9 discussions in April

- The April 7 breakout session for the Research Data Provenance Interest Group was dedicated to the proposal of a new Working Group on Provenance Patterns.
- The main goal of this group is to identify and document general provenance solution patterns, abstracted from specific best practice solutions.
- WGPP aspires to help groups like this one recognize when solutions developed in the context of a particular use case are exemplars of more general patterns.
- The pattern to which a solution conforms may have important practical implications for data integration and how easy or difficult it is to answer questions.

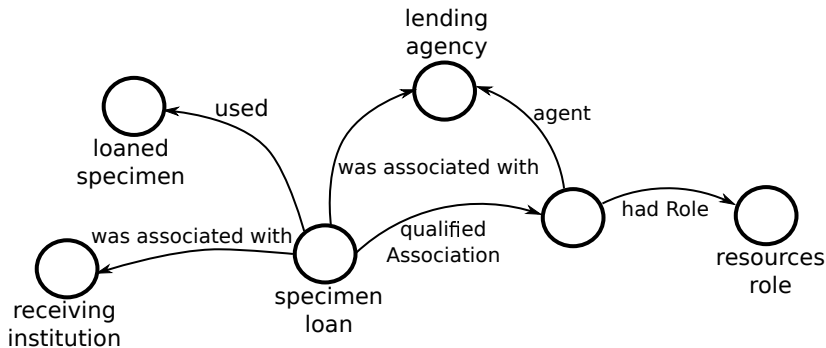
- WGPP has selected the contributor role problem as one of the first use cases for attention and analysis.

- WGPP has selected the contributor role problem as one of the first use cases for attention and analysis.
- We're preparing illustrations of how the same CRediT roles could participate in realizing three very different provenance patterns, including a PROV realization and a VIVO-ISF realization.

- WGPP has selected the contributor role problem as one of the first use cases for attention and analysis.
- We're preparing illustrations of how the same CRediT roles could participate in realizing three very different provenance patterns, including a PROV realization and a VIVO-ISF realization.
- Although those latter two solutions have identical graph structures, their differences resist direct alignment via OWL reasoning systems.

- WGPP has selected the contributor role problem as one of the first use cases for attention and analysis.
- We're preparing illustrations of how the same CRediT roles could participate in realizing three very different provenance patterns, including a PROV realization and a VIVO-ISF realization.
- Although those latter two solutions have identical graph structures, their differences resist direct alignment via OWL reasoning systems.
- Diagrammatic illustrations of the differences are presented on the following slides. The scenario is a physical specimen loan from one institution to another, where in each illustration the lending agency's **Resources Role** is from the CRediT taxonomy.

The PROV illustration



The PROV illustration

Things to notice about the PROV solution:

- 1 The CRediT role is an attribute of the *association* between the lending agency and the loan event; that's how the PROV data model works.

The PROV illustration

Things to notice about the PROV solution:

- ① The CRediT role is an attribute of the *association* between the lending agency and the loan event; that's how the PROV data model works.
- ② But since RDF admits only binary relationships, we reify the association as a vertex as well as an edge.

The PROV illustration

Things to notice about the PROV solution:

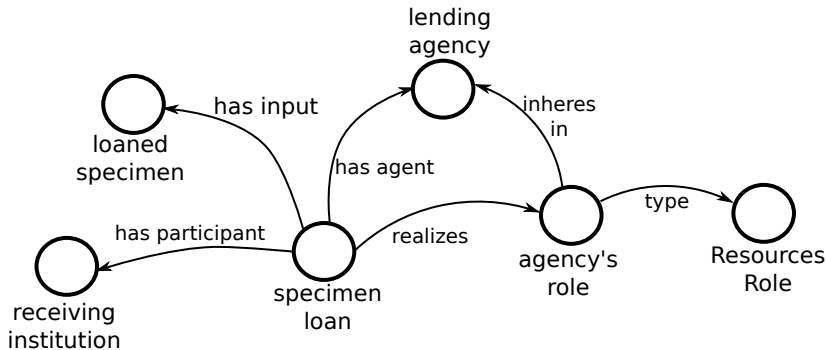
- ① The CRediT role is an attribute of the *association* between the lending agency and the loan event; that's how the PROV data model works.
- ② But since RDF admits only binary relationships, we reify the association as a vertex as well as an edge.
- ③ The **resources role** in this example is a *particular*, i.e., an instance of a **Role** class.

The PROV illustration

Things to notice about the PROV solution:

- ① The CRediT role is an attribute of the *association* between the lending agency and the loan event; that's how the PROV data model works.
- ② But since RDF admits only binary relationships, we reify the association as a vertex as well as an edge.
- ③ The **resources role** in this example is a *particular*, i.e., an instance of a **Role** class.
- ④ Different lending agencies could play that very same **resources role** in the context of other loan events.

The VIVO-ISF illustration



The VIVO-ISF illustration

Things to notice about the VIVO-ISF solution:

- 1 VIVO-ISF is based in BFO, the Basic Formal Ontology, which has an Aristotelian characterization of roles. Roles inhere within the agents that assume them.

The VIVO-ISF illustration

Things to notice about the VIVO-ISF solution:

- ① VIVO-ISF is based in BFO, the Basic Formal Ontology, which has an Aristotelian characterization of roles. Roles inhere within the agents that assume them.
- ② The **Resources Role** in this example is not an instance or particular, but a *class*. That is because no role instance can inhere within two different agents. Two agent roles may have the same *class identity*, but not the same *individual identity*.

The VIVO-ISF illustration

Things to notice about the VIVO-ISF solution:

- 1 VIVO-ISF is based in BFO, the Basic Formal Ontology, which has an Aristotelian characterization of roles. Roles inhere within the agents that assume them.
- 2 The **Resources Role** in this example is not an instance or particular, but a *class*. That is because no role instance can inhere within two different agents. Two agent roles may have the same *class identity*, but not the same *individual identity*.
- 3 The OWL-DL subset of OWL places restrictions on what we can assert and deduce about classes. Classes cannot be treated as individuals, and class axioms admit only a limited number of properties. One implication is that we can't align or map between these two examples using DL reasoning software.

The VIVO-ISF illustration

Things to notice about the VIVO-ISF solution:

- ① VIVO-ISF is based in BFO, the Basic Formal Ontology, which has an Aristotelian characterization of roles. Roles inhere within the agents that assume them.
- ② The **Resources Role** in this example is not an instance or particular, but a *class*. That is because no role instance can inhere within two different agents. Two agent roles may have the same *class identity*, but not the same *individual identity*.
- ③ The OWL-DL subset of OWL places restrictions on what we can assert and deduce about classes. Classes cannot be treated as individuals, and class axioms admit only a limited number of properties. One implication is that we can't align or map between these two examples using DL reasoning software.
- ④ The alignment can be accomplished using ordinary programming languages, but then our mapping and integration rules will be much harder to verify and maintain.

A third pattern: roles as contingent agent subclasses

- The last example illustrates a third way we can represent contributor roles.

A third pattern: roles as contingent agent subclasses

- The last example illustrates a third way we can represent contributor roles.
- Being a **Resource Provider** is understood as a class identity for the agents themselves. This class identity is deduced (dashed edge) from the agent's participation in a **Resources Provision** event.

A third pattern: roles as contingent agent subclasses

- The last example illustrates a third way we can represent contributor roles.
- Being a **Resource Provider** is understood as a class identity for the agents themselves. This class identity is deduced (dashed edge) from the agent's participation in a **Resources Provision** event.
- The participation that licenses this inference is not directly associated with the triggering event (as it was in the realization of the BFO role). If we wish to learn why the lending agency is classified as a **Resource Provider** we must rely on the reasoner that made the deduction to provide an explanation.

The contingent subclass pattern

